Machine learning methods for binary and multiclass classification of melanoma thickness from dermoscopic images

Aurora Sáez, *Student Member, IEEE,* Javier Sánchez-Monedero, Pedro Antonio Gutiérrez, *Senior Member, IEEE,* and César Hervás-Martínez, *Senior Member, IEEE*

***Abstract*—Thickness of the melanoma is the most important factor associated with survival in patients with melanoma. It is most commonly reported as a measurement of depth given in millimeters (mm) and computed by means of pathological examination after a biopsy of the suspected lesion. In order to avoid the use of an invasive method in the estimation of the thickness of melanoma before surgery, we propose a computational image analysis system from dermoscopic images. The proposed feature extraction is based on the clinical findings that correlate certain characteristics present in dermoscopic images and tumor depth. Two supervised classification schemes are proposed: a binary classification in which melanomas are classified into thin or thick, and a three-class scheme (thin, intermediate, and thick). The performance of several nominal classification methods, including a recent interpretable method combining logistic regression with artificial neural networks (Logistic regression using Initial variables and Product Units, LIPU), is compared. For the three-class problem, a set of ordinal classification methods (considering ordering relation between the three classes) is included. For the binary case, LIPU outperforms all the other methods with an accuracy of 77.6%, while, for the second scheme, although LIPU reports the highest overall accuracy, the ordinal classification methods achieve a better balance between the performances of all classes.**

***Index Terms*—Melanoma thickness, dermoscopic image, machine learning, ordinal classification, nominal classification**

I. Introduction

MELANOMA is a malignancy of melanocytes, the cells that produce the pigment melanin that colors the skin, hair and eyes. Melanoma that occurs on the skin, called cutaneous melanoma, is the most common type of melanoma. It is well accepted that only early detection can reduce mortality, since the prognosis of patients with melanoma depends on the thickness of the tumor at the time of surgical treatment [1].

If the melanoma is confined to the epidermis, it is an in situ melanoma, curable by adequate removal with surgery. When the cancerous cells have grown through the basement membrane into the deeper layer of the skin (the dermis), it is known as invasive melanoma, whose prognosis worsens with depth of invasion.

Breslow index [2] is a method to measure the depth of melanoma invasion by means of pathological examination after incisional or excisional biopsy of the suspected lesion [3]. It is measured vertically in millimeters from the top of the granular layer of the epidermis to its deepest part within the dermis. It is a valuable tool in prognosing patients survival [3]. Moreover, it represents the main parameter used to establish the width of surgical margins of excision [4] [5], as well as to select patient for sentinel lymph node biopsy (SNB) [4] [6]. SNB is a surgical procedure used to determine if cancer has spread beyond a primary tumor into the lymphatic system. Therefore, measuring the thickness of the melanoma before surgical excision is crucial to assess the high or low risk of progression, to ensure adequate excision margins avoiding a second more radical operation and to perform SNB if needed.

According to melanoma thickness, it can be classified as indicated in Table I.

TABLE I
STAGES OF MELANOMA ACCORDING TO THICKNESS

Stage I (thin)	$<$0.76 mm
Stage II (intermediate)	0.76 mm - 1.5 mm
Stage III (thick)	$>$1.5 mm

However, Breslow index can be erroneous determined if the section of the excised tumor is not made along the thickest part of the tumor. Moreover, taking into account that melanoma has the potential to be diagnosed through non-invasive approaches because of its cutaneous location [7], some authors [3], [8] suggest that determination of Breslow index by non-invasive techniques, such as dermoscopy or ultrasonography, would be a great advance in every day clinical management of melanoma.

Dermoscopy is a non-invasive technique that uses light and magnification that allows *in vivo* visualization of morphologic structures in pigmented lesions correlated with specific histologic architectural characteristics not otherwise visible to the unaided eye, increasing melanoma diagnostic accuracy by up to 35% [9], [10].

This technique has already given promising results in Breslow index determination [3]. Some authors have determined a possible relation between tumor thickness and scores obtained with diagnosis methods from dermoscopic images, such as

A. Sáez is with the Signal Theory and Communications Department, University of Seville, 41092 Seville, Spain (e-mail: aurorasaez@us.es).

J. Sánchez-Monedero is with Department of Mathematics and Engineering, Universidad Loyola Andalucía, Spain (e-mail: jsanchez@uloyola.es)

P. A. Gutiérrez and C. Hervás-Martínez are with Department of Computer Science and Numerical Analysis, University of Córdoba, 14004 Córdoba, Spain (e-mail: pagutierrez@uco.es; chervas@uco.es).

ABCD [11] or seven point checklist [12]. However, most of the works have studied the relation between dermoscopic structures, color features, or even features of shape and the depth of melanoma. In this sense, Argenziano et al. [13] indicated that pigment network, gray-blue areas, atypical vascular pattern, and diameter of more than 1.5 mm allowed prediction of thickness when melanomas were categorized in two groups: less than 0.76 mm, including in situ (thin) melanomas, and more or equal than 0.76 mm (thick melanomas). Stante et al. [4] fully confirmed the connection between these dermoscopic criteria and histological thickness of melanoma. In [14], the relationship between the observed dermatoscopic findings and the histopathological thickness was also studied, concluding that pigment network, light-brown color, and irregularity are more frequent in thin melanomas ($<$0.76 mm) and these findings gradually disappear and are replaced by vascular pattern, gray-blue areas, and white scar-like areas in thick melanomas ($\geq$0.76 mm). In [15], Argenziano et al. demonstrated that the presence of a combination of blue and black color within the lesion is a simple dermoscopic clue to recognize pigmented nodular melanoma, which presents a high thickness. More recently, Silva et al. [16] described the most frequent criteria in thin melanomas: asymmetry in two axes, 3 or more colors, atypical dots or globules, and pigment network.

In recent decades, computerized dermoscopy image analysis systems have been proposed to assist the diagnosis of pigmented lesions [17]. Most of the works found in the literature have focused on distinguishing melanomas from benign lesions [18], [19]. However, the problem of characterization of the different types of melanoma according its thickness by image analysis poses a great challenge for its automatic implementation. To the best of our knowledge, only a work [20] has addressed this problem, classifying tumors into two classes: thin and thick melanomas. The authors classified the melanoma depth by a commercial software using 49 features related to color, geometry, and texture integrated therein. 141 images from a private database, extracted by the hardware system property of the own company, were used.

In this work, we propose a computational image analysis system to estimate the thickness of melanoma from dermoscopic images based on above mentioned the clinical findings. The dermoscopic images were extracted from the Interactive Atlas of Dermoscopy, published by Edra Medical Publishing New Media [21]. A feature extraction is performed inspired by the clinical findings, selecting a set of features correlated with melanoma thickness. The main motivation of the proposal is to classify melanoma thickness without using invasive methods. Considering that melanoma thickness is a continuous variable correlated with prognosis (and thus with patient survival), different classification options can be explored. We first explore the performance in the binary case, and then we address the challenge of a finer patient prognosis task with three stages of depth. In this way, the objective of the binary classification is to distinguish between melanomas $<$0.76 mm, including in situ ones, and those with $\geq$0.76 mm of thickness. The three-class scheme, which, to the authors knowledge, has not been previously studied in the literature, approaches the classification of melanomas considering three stages of depth: $<$0.76 mm, 0.76 mm – 1.5 mm and $>$1.5 mm, i.e. thin, intermediate, and thick, respectively, according the classification proposed in [13].

The classification performance of several machine learning methods is compared, considering three performance metrics and the interpretation possibilities of the obtained models. A recent model combining logistic regression with artificial neural networks (Logistic regression using Initial variables and Product Units, LIPU) is applied to the problem. Its main advantage is that it leads to interpretable probabilistic models, maintaining a considerable level of accuracy. Moreover, we select a set of ordinal classification methods for the three-class problem, which are machine learning algorithms that achieve better performance in multi-class tasks when there is an ordering relationship between the classes, gaining recent attention [22], [23]. Our results show that LIPU model obtains accurate results both for the binary and three-class versions of the problem, while ordinal classification methods achieve a better balance between the accuracies obtained for all classes in the three-class problem, this dealing to more robust behaviour in patient diagnosis.

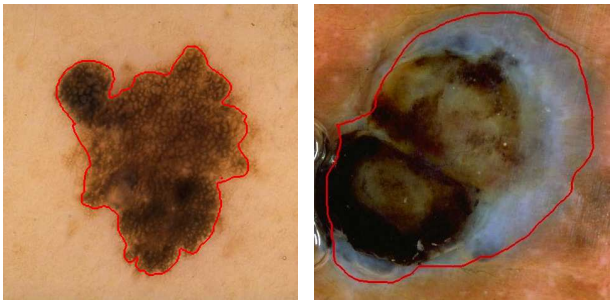
II. Database

The image database used is constituted by 250 dermoscopic images of melanomas: 167 melanomas $<$0.76 mm, including 64 in situ melanomas, 54 melanomas 0.76-1.5 mm and 29 melanomas $>$1.5 mm. As clinical experts in melanoma diagnosis have reported in their works [4], [13], melanomas in situ are included in the first class, because they are considered thin melanomas. Moreover, a specific study of these lesions [1] concluded that dermoscopic criteria for in situ melanomas are similar to those for $<$0.76 mm of thickness. All images were extracted from the Interactive Atlas of Dermoscopy, published by Edra Medical Publishing New Media [21], which is a multimedia project for medical education with images of pigmented skin lesions from different hospitals. All lesions were biopsied and diagnosed histopathologically.

The images, which have the same resolution (768x512 pixels), were segmented using the automatic segmentation algorithm proposed in [24], in which an edge based level-set technique is applied together a perceptually adapted color gradient [25]. Some examples of segmented melanomas are shown in Fig. 1. An image normalization process was not performed due to lack of information about the procedure of image acquisition and camera calibration. However, a selection of images was carried out discarding, those in which the lesion did not fit entirely within the image frame and presented too many artifact types.

III. Feature extraction

In this section, the features used to characterize the melanomas are described. The extraction is inspired by the findings derived from clinical studies regarding the correlation between certain characteristics seen in dermoscopic images and melanoma thickness. These findings mainly involve dermoscopic structures and color features. A total of 81 descriptors (x_1-x_{81}) based on shape, color and texture are extracted.



(a) (b)

Fig. 1. Examples of segmented melanomas: (a) melanoma <0.76 mm, (b) melanoma ≥0.76 mm

A. Shape features

Argenziano et al. [13] indicated that a diameter of more than 15 mm was more frequent in melanomas with >0.76 mm of thickness, and Silva et al. [16] that one of the most frequent criterion in thin melanomas was asymmetry in two axes. To calculate these criteria, we compute area (x_1), as number of pixels inside the lesion border, eccentricity (x_2), as circularity measure, perimeter (x_3), and major axis of the lesion (x_4). Eccentricity is computed as $\sqrt{1-\frac{b^2}{a^2}}$, were a and b are the major and the minor axes of the ellipse that has the same second-moments as the region resulting from the lesion segmentation. A ellipse whose eccentricity is zero is a circle. Eccentricity can be considered a special case of asymmetry [14].

B. Color features

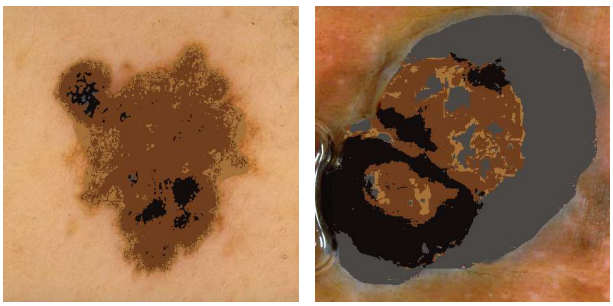
The main dermoscopic structures which are more discriminative in the melanoma thickness estimation are: pigment network, blue-gray veil, vascular pattern and white scar-like areas. These structures are associated with different colors: pigment network is associated with black and brown, vascular pattern with red color, as the name suggests, blue-gray veil with blue and gray colors, and white scar-like areas with white. Moreover, Lorentzen et al. [14] indicated that light brown color was more frequent in thin melanomas (<0.76 mm), as well as Argenziano et al. [15] established a relationship between a combination of blue and black color within the lesion with melanomas with high thickness. Therefore, color features play an important role in the estimation of depth of the melanoma.

We compute two sets of color features: the first one related to the six colors assessed in the pigmented lesions, and the second one related to statistics computed over the color of the whole lesion. Regarding the first color feature set, it is important to note that there are six main colors that a pigmented lesion can present: black, dark brown, light brown, blue-gray, red, and white [26]. These colors seen in dermoscopy depend to some extent on the depth of the lesion, since the different colors appear depending on how deeply in the skin the melanina is located [27]: melanin appears black when it is located in the stratum corneum and upper epidermis. Deeper in the epidermis, it appears brown. In the dermis, melanin appears either gray or blue. Red is associated with dilation of blood vessels and white with regression and/or scaring. Thus, we propose to segment each lesion into their constituting colors.

In order to achieve this, a similar approach to that proposed by Seidenari et al. [28] was followed. We developed a color palette comprising the six color groups. The palette is formed by 144 patches or samples (24 patches per color), 40×40 pixels, manually extracted by an expert in pigmented lesion diagnosis from 60 lesion images, different from the 250 melanoma images used for our study. Each patch presents unequivocally one of the six possible colors (white, blue-gray, dark brown, light brown, black and red). Fig. 2 shows some examples of patches belonging to each color class.

Fig. 2. Sample of the color palette used for color group attribution. The colored square at the beginning of the column corresponds to the false color attributed to the pixels belonging to that color group.

The palette was used to extract the color regions of the lesions from the patches according to a nearest neighbor approach. In [28], each pixel of the image was assigned to the color patch that minimized its Euclidean distance in the RGB color space. In this paper, CIE $L^*a^*b^*$ color space, developed by International Commission on Illumination (CIE), is used, because, unlike RGB, it is a approximately perceptually uniform color space. In Fig. 3, it can be seen the color area identification in the melanomas shown in Fig. 1.



(a) (b)

Fig. 3. Example of color area identification in the melanomas shown in Fig. 1. a) <0.76 mm, b) ≥0.76 mm

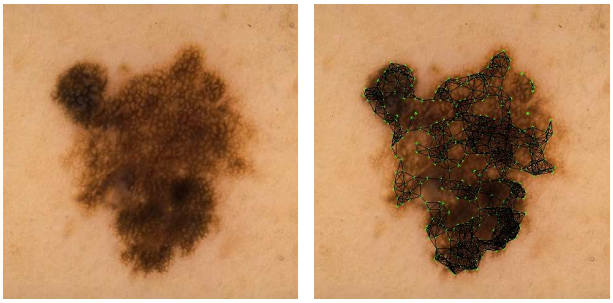
From this color identification, seven features are extracted. Six descriptors (x_5-x_{10}) that represent the percentage of the lesion area classified as these colors. And one more (x_{11}) that represents the number of colors that each lesion presents, criterion established by Silva et al. [16], taking into account in this case that a color must cover at least 1% of the lesion to be counted.

Moreover, color of whole lesions were quantified by four statistics (mean, standard deviation, SD, kurtosis, Ku, and skewness, Sk) over the three channels of two color spaces: RGB and CIE $L^*a^*b^*$, i.e. R, G, B, L^*, a^*, and b^* (obtaining 24 additional color features, x_{12}-x_{35}).

C. Pigment network features

Many authors have found a relation between the depth of melanoma and certain dermoscopic structures. Among them, the pigment network seems to be the most cited discriminative feature [4], [13], [14], [16]. Its occurrence inversely correlates with melanoma thickness [21]. Due to its importance, we propose to extract specific features derived from pigment network detection.

A pigment network is defined as a regular grid of brownish lines over a diffuse light-brown background [21]. Our aim is to seek the 'holes' of the network. The first step is to apply a top hat filter over the channel of lightness (L^*) of CIE $L^*a^*b^*$ color space. Top hat transform extracts objects brighter than their surroundings. After a thresholding by Otsu's method [29], a binary image with areas that could possibly belong to a pigment network is obtained. In order to remove those wrongly detected areas, we apply the two conditions relative to area size and color proposed in the work presented by Sadeghi et al. [30], which uses images extracted from the same Atlas [21]. To visualize the detected pigment network, a graph, whose nodes are centers of the detected holes belonging to the pigment network, is created. Nodes within a maximum distance threshold, set to 2.5 times the average diameter of holes, are connected together. Fig. 4 show an example of pigment network detection.



(a) (b)

Fig. 4. (a) Original image (b) Pigment network detection

To decide if a lesion presents pigment network, the density ratio proposed in [30] is computed:

$$Density = \frac{|E|}{|V|\log(LesionSize)}, \quad (1)$$

where E is the number of edges in the graph, V is the number of nodes of the graph, and $LesionSize$ is the size of the area of the lesion, as number of pixels, used to normalize the ratio $|E|/|V|$.

From this pigment network detection, 3 features are extracted: density ratio (x_{36}), number of nodes (x_{37}) and number of links or edges (x_{38}).

D. Texture features

Besides pigment network, other dermoscopic structures have been found to have relation with the depth of melanoma, such as vascular pattern [4], [13] , blue-gray veil [4], [13], white scar-like areas [14], and dots or globules [16]. These structures are usually associated with texture features. For example, vascular pattern is associated to the presence of a vascular vessel with line shape, and gray-blue areas and white scar-like areas are found as homogeneous areas [14].

In attempt to identify different structures present in a lesion, we propose to extract three sets of texture features from three different approaches, detailed below.

1) Gray level co-occurrence matrix (GLCM): The gray level co-occurrence matrix is considered, given its effectiveness in pigmented lesion classification [19], [31], [32]. It counts how often a pixel with gray intensity of i occurs adjacent to a pixel with gray intensity of j. Therefore, it is necessary to convert the original RGB image into a grayscale image. We select the color channel of the RGB image with the highest entropy [31]. The images are uniformly quantized to 64 gray levels, as previously proposed [19], and the GLCM is computed for four different orientations of $\{0°, 45°, 90°, 135°\}$. Finally, the statistics calculated from these matrices were averaged. For each orientation, 19 statistical texture descriptors were extracted and then averaged (x_{39}-x_{57}); 12 proposed by Haralick et al. [33]: Contrast, Correlation, Entropy, Variance, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation 1, Information measure of correlation 2, and Inverse difference moment; 4 proposed in [34]: Autocorrelation, Dissimilarity, Energy, and Maximum probability; and 3 proposed in [35]: Inverse difference, Inverse difference normalized, Inverse difference moment normalized.

2) Markov random fields (MRF): A model based on Markov random fields is used to extract texture features that allow to identify different dermoscopic structures, as was proposed in [24]. These models assume that the intensity at each pixel in the image depends on the intensities of the neighboring pixels. The idea is to model a dermoscopic image in $L^*a^*b^*$ color space, and to treat the estimated parameters of the model as texture features. The texture features are constituted by the following parameters: $\mu_s, \widehat{\sigma}_s^2, \widehat{\theta}_{s,t} : t \in \eta_g$, where $\eta_g = \{t_1, t_2, t_3, t_4\} = \{(0,1),(1,0),(1,1),(-1,1)\}$ is a set of shift vectors, μ_s is the mean of the color pixels, $\widehat{\sigma}_s^2$ is the estimation of the noise variance, and the other four components, $\widehat{\theta}_{s,t}$, are the estimation of the correlation coefficients. As these features are computed from the three components of the $L^*a^*b^*$ color space, finally, we obtain 18 parameters (x_{58}-x_{75}). These parameters are estimated by the least-squares method proposed by Manjunath and Chellappa [36] (see [24] for details of the model and parameter calculation).

3) Local binary pattern (LBP): Just as GLCM, the local binary pattern (LBP) histograms [37] have been widely applied to the characterization of texture in pigmented lesions [38], [39]. This technique labels the pixels of an image by thresholding the neighborhood of each pixel based on its value (neighbor pixels with value higher or equal to the center pixel value assume 1 and 0 otherwise) and generates a number to

quantify the local texture. A histogram is used to describe the texture information of the whole lesion. We compute LBP features over the same channel that was computed the GLCM, using eight sampling points on a circle of radius $R = 2$ and $R = 10$ pixels (see [37] for more details). Three statistics (SD, Ku and Sk) over the two LBP histograms are extracted (x_{76}-x_{81}).

IV. Classification methods

Our purpose is to study the classification performance in two variants of the dataset. The binary case try to distinguish between thin melanomas, including in situ ones, and the rest of melanomas (Stage I vs. Stages II and III, see Table I), having 167 and 83 patterns of each class respectively. A second approach to the melanoma thickness detection is proposed in a three-class scheme: class one ($\mathcal{C}_1$) corresponds to thin melanoma; class two ($\mathcal{C}_2$) is the intermediate melanoma; and class three ($\mathcal{C}_3$) are thick melanomas (see Table I). The classes distribution here is 167, 54 and 29. Since we can assume a natural ordering between melanoma types ($\mathcal{C}_1 \prec \mathcal{C}_2 \prec \mathcal{C}_3$), for the second approach we select a set of ordinal classification methods (see Section IV-B). In addition to the data ordering, the dataset is an imbalanced datataset: this is, the distribution of patterns per class varies significantly so that learning algorithms tend to bias the model performance towards the majority classes [40]. For this reason we include several performance metrics [40], [41] to perform a more robust evaluation (see Section V-A).

A. Logistic regression using Initial variables and Product Units (LIPU)

Let $D = \{(\mathbf{x}_n, \mathbf{y}_n); n = 1, 2, ..., N\}$ be a training dataset, where $\mathbf{x}_n = (x_{1n}, ..., x_{kn})$ is the vector of input variables (feature vector) taking values in $\mathbf{\Omega} \subset \mathbb{R}^k$, and $\mathbf{y}_n$ is the label of the n-th individual, where a "1-of-J", J is the number of classes, encoding vector is used, $\mathbf{y} = \left(y^{(1)}, y^{(2)}, ..., y^{(J)}\right)$, such that $y^{(l)} = 1$ if $\mathbf{x}$ corresponds to an image belonging to class $\mathcal{C}_l$, and $y^{(l)} = 0$ otherwise.

The LIPU model [42], [43] is a combination of a logistic regression model with basis functions obtained from Product Unit Neural Networks (PUNN). The basis function of the hidden neurons of PUNN is the Product Unit (PU) function, where the output of the neuron is the product of their inputs raised to real valued weights. They are an alternative to sigmoidal neural networks, based on multiplicative nodes instead of additive ones. They have the ability to express strong interactions between input variables, providing big variations at the output from small variations at the inputs [44]. In this way, the predictor function includes a standard linear model in the input variables and a nonlinear model constructed with PU variables, which captures interactions in the input space. The general expression of the predictor function is given by:

$$f_l(\mathbf{x}, \boldsymbol{\theta}_l) = \beta_0^l + \sum_{i=1}^{k} \beta_i^l x_i + \sum_{j=k+1}^{k+m} \left(\beta_j^l \left(\prod_{i=1}^{k} x_i^{w_{ji}} \right) \right), \quad (2)$$

where $l = 1, 2, ..., J-1$, k is number of input variables, m is number of PU nonlinear transformations, $\boldsymbol{\theta}_l = (\boldsymbol{\beta}^l, \mathbf{W})$ is the vector of parameters for each predictor function, $\boldsymbol{\beta}^l = \left(\beta_0^l, \beta_1^l, ..., \beta_{k+m}^l\right)$ are the coefficients of the multilogistic regression model, $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m)$ are the parameters of the PU nonlinear transformations, and $\mathbf{w}_j = (w_{j1}, ..., w_{jk})$ includes the parameters of the j-th PU transformation, where each weight w_{ji} is an exponent applied to the i-th input value. As can be observed, the model corresponds to a standard logistic regression model extended with PU transformations, where each PU is $B_j(\mathbf{x}, \mathbf{w}_j) = \left(\prod_{i=1}^{k} x_i^{w_{ji}}\right)$.

To estimate the parameters $\boldsymbol{\theta}$, one possibility is the minimization of the negative log-likelihood function $L(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=1}^{N} \left[-\sum_{l=1}^{J-1} y_n^{(l)} f_l(\mathbf{x}_n, \boldsymbol{\theta}_l) + \log \sum_{l=1}^{J-1} \exp f_l(\mathbf{x}_n, \boldsymbol{\theta}_l) \right]$, where $\boldsymbol{\theta}_l = (\boldsymbol{\beta}^l, \mathbf{W})$ and $f_l(\mathbf{x}_n, \boldsymbol{\theta}_l)$ corresponds to the LIPU model defined in (2). The use of gradient-based methods to maximize this function is not recommended, because of the nonlinearity of the model with respect to the parameters w_{ji} of $\mathbf{W}$ and the indefinite character of the associated Hessian matrix of $L(\boldsymbol{\theta})$. Moreover, the optimal number of PU transformations of the model (m) is unknown. Thus, the estimation of the vector parameter $\hat{\boldsymbol{\theta}}$ is carried out by means of the hybrid procedure described below.

The methodology considered [45] is based on the combination of a global explorer (an Evolutionary Programming, EP, algorithm), and a local exploiter (maximum likelihood optimization). Firstly, the EP algorithm designs the structure and trains the weights of a PUNN, obtaining the number of PUs in the model, m, and the corresponding sparse weight matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m)$. The details of the evolutionary operators used can be consulted in [45]: a parametric mutation applies changes in the weights, and a structural mutation adds or removes hidden nodes or connections. The nonlinear transformations given by the PUs of the best PUNN in the final generation of the EP algorithm are used to extend the input space and apply a maximum likelihood optimization.

Secondly, we consider that the classification problem can be now linearly solved in the space formed by these new variables and the initial ones [45]. The remaining coefficient vector $\boldsymbol{\beta}$ is calculated by maximum likelihood optimization. We use the the *SimpleLogistic* method [46], which builds multinomial logistic regression models by using the LogitBoost algorithm. The algorithm uses simple regression functions (of only one variable) for a committee of functions, in a stagewise fitting. Using simple regression instead of multiple ones basically results in slowing down the process, but, if stopped before convergence, an automatic variable selection is done. In this way, the *SimpleLogistic* algorithm is based on applying LogitBoost with simple regression functions and determining the optimum number of iterations by cross-validation. Further details about the algorithm can be found in [46]. The motivation behind using *SimpleLogistic* is that it is able to eliminate those variables which do not contribute too much to the classification model or those which cause overfitting, simplifying the final equation of the model and allowing better interpretation.

In this way, the experiments of this paper justify the use of the hybrid LIPU model (with both logistic regression and PU functions) comparing its performance to the its two

main components: the original PUNN[1] and the SimpleLogistic algorithm (logistic regression, LR)[2].

B. Ordinal classification

Ordinal classification, also known as ordinal regression, deals with classification problems in which there is an ordering between classes. Though these problems can be tackled with nominal classifiers, specific methods have been developed to take advantage of the ordering information in the label space and to minimize the magnitude of the errors in the prediction phase [48]. For example, in our three-class problem, wrongly classifying a thick melanoma as a thin one is an error that should be penalized more than misclassifying it as an intermediate one.

For comparing the performance of nominal methods (not considering ordering relation) and ordinal ones, the battery of experiments includes five nominal methods and three ordinal ones. The three methods presented in previous subsections are nominal (LIPU, PUNN and LR) and we complement this set with Kernel Discriminat Analysis (KDA) and Support Vector Machine for Classification (SVC) [49]. For the ordinal methods, we have chosen Support Vector Ordinal Regression with implicit constraints (SVORIM) [50], RED-SVM [22], which applies the reduction from cost-sensitive ordinal ranking to weighted binary classification (RED) framework to SVM, and Kernel Discriminant Learning for Ordinal Regression (KDLOR) [23][3]. For additional information about these methods, please refer to the provided references and to a recent survey on ordinal regression [48].

V. Experiments

A. Performance evaluation and experimental design

Performance evaluation of classification tasks is a multi-dimensional issue, and model behaviour assessment depends on the specific problem requirements. In our case, we have selected the following classification performance metrics based on the application purpose:

- Accuracy (Acc) is the rate of correctly classified patterns and represents the global performance of the classification task: $Acc = \frac{1}{N}\sum_{i=1}^{N}[\![\hat{y}_i = y_i]\!]$, where $[\![c]\!]$ is the indicator function, being equal to 1 if c is true, and to 0 otherwise. Acc values range is $[0, 1]$.
- Minimum Sensitivity (MS) is proposed to measure imbalanced multi-class models performance [40], and it is defined as the minimum per class Sensitivity (S_i): $MS = \min\{S_i;\ i = 1, \ldots, J\}$, where J is the number of classes, and S_i is the accuracy taking into account only patterns from class $\mathcal{C}_i$.
- Mean Absolute Error (MAE) is the average deviation in absolute value of the predicted class from the true class. For imbalanced datasets, this measure should be modified to consider the relative frequency of the classes, deriving in the *Average* MAE ($AMAE$) [51]: $AMAE = \frac{1}{J}\sum_{j=1}^{J} MAE_j = \frac{1}{J}\sum_{q=1}^{J}\frac{1}{n_q}\sum_{i=1}^{n_q} e(\mathbf{x}_i)$, where $e(\mathbf{x}_i) = |\mathcal{O}(y_i) - \mathcal{O}(\hat{y}_i)|$ is the distance between the true and the predicted ranks, $\mathcal{O}(\mathcal{C}_q) = q$ is the position of the q-th label, n_q is the number of patterns of class $\mathcal{C}_q$, and $AMAE$ values range from 0 to $J-1$. For the binary case, $AMAE$ can be seen as a Weighted Accuracy ($WAcc$), because the errors ($e(\mathbf{x}_i)$) will be 0 or 1.

Experiments are performed using a 10-fold cross-validation. All the features were properly standardized using mean and variance of the training data of each fold. The parameter values for LIPU and PUNN algorithms can be found in [45]. To adjust the hyper-parameters of the rest of methods, a nested cross-validation is applied to the training data, with a grid search for the different values. The criteria for selecting the parameters is $AMAE$ or $WAcc$. Hyper-parameter values considered are the same as those used in [48].

B. Experimental results

Experimental results are presented in terms of generalization performance of the models in Table II. The different metrics are obtained from the sum of the 10 unseen predictions done by each method in the 10-fold setup. We provide a public website with the dataset and 10 partitions together with links to implementations of all the methods used in the experiments, as well as supplementary matherial[4]. In addition, Table III shows generalization confusion matrices for LIPU and KDA/KDLOR to complement performance analysis. The matrices for the remaining methods are included as supplementary data in the website associated to this paper. For the binary case, the LIPU has the best performance in Acc and $AMAE$, and the second best performance in MS. For the ordinal classification problem, LIPU has the best Acc; however, the performance in MS and $AMAE$ decays. On the other hand, KDLOR has the best performance in MS and $AMAE$ for the ordinal problem, achieving a 55.2% of accuracy for the worst classified class ($\mathcal{C}_3$), with a mean error magnitude below 0.5 of distance (in number of categories) between the true and predicted class. Note the performance of the LR and PUNN is improved by LIPU model in all the cases. Finally, observe that the nominal methods have the worst performance if we consider $AMAE$.

VI. Interpretability of the LIPU model

Since the LIPU methodology proposed presented the best results in Acc, it seems logical to analyze one of the models obtained using this methodology in order to understand better the cause and effect relationship between the features used as inputs and the probability estimated by the model. We focus on the binary case.

According to the model obtained for the first fold, which is selected because of its interpretability, the logit of the

[1]LIPU and PUNN source code is available at http://www.uco.es/grupos/ayrna/en/partitions-and-datasets/#paguitierrez2011ieeetnn

[2]For *SimpleLogistic*, we used the implementation available at Weka [47]

[3]For SVC, SVORIM, RED-SVM and KDLOR we have used the ORCA tool https://github.com/ayrna/orca

[4]http://www.uco.es/grupos/ayrna/ieeetmi2015

TABLE II
EXPERIMENTAL RESULTS IN MEAN CONSIDERING THREE CLASSIFICATION PERFORMANCE METRICS.

Binary problem			
Method	Acc	MS	$WAcc$
LIPU	**0.776**	*0.602*	**0.268**
LR	0.752	0.530	0.304
PUNN	0.720	0.494	0.337
KDA	0.712	**0.663**	0.300
SVC	*0.764*	0.518	*0.298*
Three-class problem			
Method	Acc	MS	$AMAE$
LIPU	**0.684**	0.185	0.656
LR	0.632	0.069	0.813
PUNN	0.648	0.148	0.759
KDLOR	0.644	**0.552**	**0.446**
SVC	*0.664*	0.259	0.675
RED-SVM	0.624	*0.345*	0.583
SVORIM	0.636	*0.345*	*0.579*

TABLE III
CONFUSION MATRICES FOR THE BEST MODELS: LIPU AND KDA/KDLOR IN THE BINARY AND ORDINAL DATABASES.

Binary problem						
	LIPU			KDA		
	insitu	insitu		insitu	insitu
rest	144	23	rest	123	44
rest	33	50	rest	28	55
Three-class problem						
	LIPU			KDLOR		
	thin	interm.	thick		thin	interm.	thick
thin	154	12	1	thin	115	47	5
interm.	40	10	4	interm.	19	30	5
thick	9	13	7	thick	3	10	16

probability that an image $\mathbf{x}$ belongs to $\mathcal{C}_1$, $\text{logit}(p(\mathbf{x} \in \mathcal{C}_1))$, is:

$$f(\mathbf{x}|\boldsymbol{\beta}, \mathbf{W}) = ln\left(\frac{p(\mathbf{x}\in\mathcal{C}_1)}{1-p(\mathbf{x}\in\mathcal{C}_1)}\right) = 14.12 + 1.308x_9 + \quad (3)$$
$$+3.346x_{17} - 4.040x_{23} - 4.981x_{45} - 2.006x_{52} -$$
$$-3.54x_{77} + 4.248B_1(\mathbf{x}, \mathbf{w_1}) - 3.219B_2(\mathbf{x}, \mathbf{w_2})$$

being:

$$\begin{aligned} B_1(\mathbf{x}, \mathbf{w_1}) &= (x_6)^{-5.26}(x_9)^{2.83}(x_{13})^{5.9}(x_{20})^{2.15} \\ &(x_{21})^{-1.49}(x_{22})^{-6.48}(x_{24})^{-0.47}(x_{26})^{-1.13}(x_{32})^{4.83} \\ &(x_{35})^{-2.31}(x_{38})^{3.80}(x_{50})^{1.11}(x_{51})^{-4.33}(x_{53})^{2.44} \\ &(x_{54})^{0.89}(x_{55})^{-3.15}(x_{56})^{-0.81}(x_{59})^{-3.47}(x_{73})^{0.40} \\ &(x_{76})^{-2.47}, \end{aligned} \quad (4)$$

$$\begin{aligned} B_2(\mathbf{x}, \mathbf{w_2}) &= (x_2)^{-5.43}(x_{14})^{-3.62}(x_{15})^{2.95}(x_{19})^{-0.75} \\ &(x_{20})^{3.63}(x_{21})^{-2.99}(x_{22})^{-1.62}(x_{24})^{-0.51}(x_{26})^{0.55} \\ &(x_{27})^{1.21}(x_{34})^{1.99}(x_{38})^{-3.03}(x_{39})^{3.25}(x_{42})^{-5.45} \\ &(x_{43})^{3.91}(x_{44})^{-0.03}(x_{45})^{1.32}(x_{46})^{0.64}(x_{48})^{3.91} \\ &(x_{49})^{1.08}(x_{58})^{1.60}(x_{61})^{2.20}(x_{64})^{-1.31}(x_{65})^{-4.08} \\ &(x_{66})^{3.57}(x_{68})^{-1.87}(x_{69})^{-3.66}(x_{70})^{-3.70}(x_{71})^{-2.19} \\ &(x_{74})^{4.16}, \end{aligned} \quad (5)$$

where the variables x_i are normalized in the range [1, 2]. For the sake of clarity, all the variables of the problem are included in Table IV, marking those used in the linear part of the model (I) or in the nonlinear one (B_1 or B_2).

Regarding the performance metrics, this model obtained values of $Acc = 0.800$, $MS = 0.625$ and $AMAE = 0.225$ for the test set. It should be noted that the model of the tenth fold resulted in values of $Acc = 0.840$, $MS = 0.775$ and $AMAE = 0.174$.

The sign of the $\boldsymbol{\beta}$ coefficients in (3) indicates the direction of the change in the probability; positive coefficients increase the probability of belonging to $\mathcal{C}_1$. In this way, the following variables increase the probability of belonging to $\mathcal{C}_1$ when their value is increased: area of black, mean b^* and the B_1 basis function. While the inverse effect (increase the probability associated to $\mathcal{C}_2$) is observed for: SD b^*, homogeneity, difference entropy, kurtosis LBP2 and the B_2 basis function.

We analyze the effect on the probability for each variable considering the minimum value of each variable (1 for the normalized variable) and setting all the other variables to 0 (i.e., eliminating its influence). The estimated probability is:

$$p(\mathbf{x} \in C_1|\boldsymbol{\beta}, \mathbf{W}) = \frac{1}{1 + e^{-f(\mathbf{x}|\boldsymbol{\beta}, \mathbf{W})}}, \quad (6)$$

The odds ratio is defined as:

$$\text{odds}(x_i) = \left(\frac{p(\mathbf{x} \in \mathcal{C}_1)}{1 - p(\mathbf{x} \in \mathcal{C}_1)}\middle| x_i = 1, x_j = 0, \forall j \neq i\right). \quad (7)$$

When the odds ratio is higher than 1 indicate, the variable influences positively in the probability of belonging to $\mathcal{C}_1$; while those lower than 1 indicate a negative effect. Thus, according to logit p terms in (3), the values of the estimated probability and the odds ratio are: for x_9 (black area), $p = 0.787$ and $\text{odds} = 2.51$; for x_{17} (mean b^*), $p = 0.966$ and $\text{odds} = 28.41$; for x_{23}, (SD b^*), $p = 0.017$ and $\text{odds} = 0.018$; for x_{45} (homogeneity), $p = 0.007$ and $\text{odds} = 0.007$; for x_{52} (diff. entropy), $p = 0.119$ and $\text{odds} = 0.135$; for the first product unit ($B_1(\mathbf{x}, \mathbf{w_1})$), $p = 0.986$ and $\text{odds} = 69.92$; and for the second product unit ($B_2(\mathbf{x}, \mathbf{w_2})$), $p = 0.038$ and $\text{odds} = 0.04$.

Note that the variable with higher influence on being an image of $\mathcal{C}_1$ is x_{17} (mean of b^* component). Taking into account that negative values in the b^* component indicate blue color and positive values yellow color, a high value of the mean of b^* in a lesion means a low level of blue color. Therefore, those images with a low level of blue color have more probability of belonging to $\mathcal{C}_1$ than $\mathcal{C}_2$. What makes sense, because the blue color is associated with gray-blue areas, structure highly suggestive of melanomas with thickness $\geq$0.76 mm ($\mathcal{C}_2$). On the contrary, the variable that more increases the probability of belonging to $\mathcal{C}_2$ is x_{45} (GLCM homogeneity). This means that a lesion with a homogeneous texture has more probability of belonging to $\mathcal{C}_2$ than $\mathcal{C}_1$. This conclusion is justified since structures predominant in thick melanomas ($\mathcal{C}_2$), such as gray-blue areas and white scar-like areas, are found as homogeneous areas [14]. However, structures such as pigment network or globules, predominant in thin melanomas ($\mathcal{C}_1$), present lower values in GLCM homogeneity, although they can show repetitive structures.

Moreover, the interactions between some features were considerably important. This fact can be observed by the analysis of the two basis functions, B_1 and B_2. Regarding B_1, we can say that the blue component of the RGB color space, the number of links in a pigment network and several texture features extracted from the GLCM show a strong interaction. Related to B_2, features extracted from the color channels of both RGB and CIE $L^*a^*b^*$ color spaces, number of links in a pigment network and texture features extracted from the GLCM and MRF models mainly compose its structure.

Finally, the study of the most frequent features in the linear part of the ten models (from the 10 folds) is especially important. They correspond to the area of black color (x_9), mean and standard deviation of b^* (x_{17}, x_{23}), density ratio of pigment network (x_{36}), homogeneity (x_{45}), inverse difference moment normalized (x_{57}), parameter θ_{4L^*} of the MRF model (x_{62}) and kurtosis of LBP2 histogram (x_{77}). Therefore, the set of characteristics more relevant is the set of the color features. On the contrary, the shape features do not have relevance in the linear part, but they do when interacting with others, especially the eccentricity, since it is the descriptor with the highest frequency of appearance in the PUs of all models. The remaining sets of features are also present in the lineal part with at least one variable belonging to each one. It is the case of network density inside the network feature set and three features related to the three sets of texture features.

The fact that features related to blue color, black color, and network density, associated with the dermoscopic structures of gray-blue areas and pigment network, appear in the models confirms what most of authors have shown in their works [13], [4], [14]: these structures allow prediction of thickness when melanomas are categorized in two groups.

VII. Summary and Conclusion

This paper proposes a computerized system to estimate the melanoma thickness from dermocospic images. The main motivation is to ensure the correct surgical margins that are made when the lesion is excised (by estimation of melanoma thickness), employing a non-invasive technique and thus avoiding its determination by a biopsy. The prognosis of a patient with melanoma depends on the thickness of the tumor, hence the importance in its degree estimation.

For this purpose, a supervised classification approach is proposed. The feature extraction step is inspired by the findings derived from clinical studies, which correlate certain characteristics seen in dermoscopic images with the depth of the tumor. It is worth mentioning the extraction process of certain characteristics: features related to the six colors assessed in the pigmented lesions by a segmentation of each lesion into their constituting colors; features related to the dermoscopic structure of pigment network, extracted from a pigment network detection method; and texture features extracted from a Markov random field model proposed for the global pattern detection in pigmented lesions.

The classification performance has been studied in two variants of the database. Firstly, a binary case, to distinguish between melanomas with thickness $<$0.76 mm, and those with thickness $\geq$0.76 mm, is studied. Three performance metrics (Accuracy, Acc, Minimum Sensitivity, MS, and Average Mean Absolute Error, $AMAE$) are used, and the LIPU model shows the best performance Acc and $AMAE$ in mean, obtaining a 77.6% and 0.268, respectively, and the second best performance in MS with a value of 60.2%.

A recent study of Rubegni et. al [20] presents a melanoma thickness detection system, reporting accuracy of 86.5%. However, a direct comparison to this study is not possible for two reasons. First, the dataset used in that study is private, with a smaller number of patterns and with a different threshold to distinguish between thin and thick melanomas. Secondly, the experimental design of the previous work is a leave-one-out procedure, which is, in general, more expensive in computation time. In this paper, we use a 10-fold cross-validation on a public dataset with more patterns and different thresholds for determining the melanoma stages. The accuracy of the best model is 84%. Although the two studies are not directly comparable, we can state that their performances are of the same order of magnitude. In addition, we provide a more extensive evaluation of the proposal in terms of experiments, performance metrics and complexity of the problem (we address the three-class problem).

The main advantage of LIPU method is that is an interpretable model, which also provides probabilistic classes assignment and perform feature selection during the learning phase. In this sense, a study of how the features contribute to the classification model is presented. All features were present in the 10 models, either in the linear part or interacting with others in the basis functions (non-linear part of the model). However, some of them present a greater relevance because they are included individually in the majority of the models. This is the case of features related to black and blue colors, pigment network, homogeneity and some texture features extracted from three approaches proposed. In spite of some authors findings which indicate that shape features, red color associated to vascular pattern and white color associated to white scar-like areas can be relevant, in our system, these features do not have an individual influence in the linear part, although they do interact with others on the basis functions.

Secondly, a three-class scheme is proposed, in which melanoma is classified into three depth stages. This second approach is motivated by the fact that the melanoma depth is correlated with the patients survival, and therefore, a finest estimation of tumor thickness will lead to a more accurate diagnosis. Due to complexity of the problem, not only nominal clasificacion methods are applied, but also ordinal ones, which assume the natural ordering between the melanoma types and exploit this ordering information to improve performance and reduce the magnitude of classification errors. To the best of our knowledge, it is the first time that these type of methods are applied to a pigmented lesion recognition problem. The results obtained for this case show that an ordinal method achieves a better balance between the performances obtained for all classes and reduces the magnitude of the errors. The complexity of the problem lies in the distinction between stage II and III, but the system is able to attain more than 55% of correct classification rate in average for the worst

TABLE IV
VARIABLES OF PROBLEM. THOSE INCLUDED IN THE LINEAR PART ARE MARKED WITH (I), THOSE IN THE FIRST PU WITH (B_1) AND THOSE IN THE SECOND WITH (B_2).

Name	Variable	Name	Variable	Name	Variable	Name	Variable
SHAPE		SD B	$x_{20}(B_1)(B_2)$	Contrast	x_{40}	θ_{3L^*}	$x_{61}(B_2)$
Area	x_1	SD L^*	$x_{21}(B_1)(B_2)$	Correlation	x_{41}	θ_{4L^*}	x_{62}
Eccentricity	$x_2(B_2)$	SD a^*	$x_{22}(B_1)(B_2)$	Dissimilarity	$x_{42}(B_2)$	σ^*_L	x_{63}
Perimeter	x_3	SD b^*	$x_{23}(I)$	Energy	x_{43} (B_2)	μ^*_a	$x_{64}(B_2)$
Major axis	x_4	$Ku.$ R	$x_{24}(B_1)(B_2)$	Entropy	$x_{44}(B_2)$	θ_{1a^*}	$x_{65}(B_2)$
SIX COLORS		$Ku.$ G	x_{25}	Homogeneity	$x_{45}(I)(B_2)$	θ_{2a^*}	$x_{66}(B_2)$
White area	x_5	$Ku.$ B	$x_{26}(B_1)(B_2)$	Max. Probability	$x_{46}(B_2)$	θ_{3a^*}	x_{67}
Blue area	$x_6(B_1)$	$Ku.$ L^*	$x_{27}(B_2)$	Variance	x_{47}	θ_{4a^*}	x_{68} (B_2)
Dark area	x_7	$Ku.$ a^*	x_{28}	Sum average	x_{48} (B_2)	σ^*_a	$x_{69}(B_2)$
Light area	x_8	$Ku.$ b^*	x_{29}	Sum Variance	$x_{49}(B_2)$	μ^*_b	x_{70} (B_2)
Black area	$x_9(I)(B_1)$	$Sk.$ R	x_{30}	Sum entropy	$x_{50}(B_1)$	θ_{1b^*}	$x_{71}(B_2)$
Red area	x_{10}	$Sk.$ G	x_{31}	Diff. variance	$x_{51}(B_1)$	θ_{2b^*}	x_{72}
No. colors	x_{11}	$Sk.$ B	$x_{32}(B_1)$	Diff. entropy	$x_{52}(I)$	θ_{3b^*}	$x_{73}(B_1)$
COLOR CHANNELS		$Sk.$ L^*	x_{33}	Measure Corr. 1	$x_{53}(B_1)$	θ_{4b^*}	$x_{74}(B_2)$
Mean R	x_{12}	$Sk.$ a^*	x_{34} (B_2)	Measure Corr. 2	$x_{54}(B_1)$	σ^*_b	x_{75}
Mean G	$x_{13}(B_1)$	$Sk.$ b^*	$x_{35}(B_1)$	Inv. Diff.	$x_{55}(B_1)$	LBP	
Mean B	$x_{14}(B_2)$	PIGMENT NETWORK		Inv. Diff. Norm.	$x_{56}(B_1)$	SD LBP2	$x_{76}(B_1)$
Mean L^*	$x_{15}(B_2)$	Density	x_{36}	Inv. Diff. Mom. Norm.	x_{57}	$Ku.$ LBP2	$x_{77}(I)$
Mean a^*	x_{16}	Nodes	x_{37}	MRF		$Sk.$ LBP10	x_{78}
Mean b^*	$x_{17}(I)$	Links	$x_{38}(B_1)(B_2)$	μ^*_L	$x_{58}(B_2)$	SD LBP10	x_{79}
SD R	x_{18}	GLCM		θ_{1L^*}	$x_{59}(B_1)$	$Ku.$ LBP10	x_{80}
SD G	$x_{19}(B_2)$	Autocorrelation	$x_{39}(B_2)$	θ_{2L^*}	x_{60}	$Sk.$ LBP10	x_{81}

class. Finally, in contrast to nominal classification models, the purpose of ordinal models is to reduce the magnitude of errors in such a way that, when patterns are misclassified, the label predicted is as close as possible to the real label (i.e. prediction errors are generally bounded to the neighbor classes). This results in a more reliable prognosis system.

ACKNOWLEDGMENT

This work was partly financed by a grant provided by the TIN2014-54583-C2-1-R project of the Spanish Ministry of Economy and Competitively (MINECO), by FEDER Funds and by the P11-TIC-7508 project of the Junta de Andalucía, Spain.

REFERENCES

[1] M. Pizzichetta, G. Argenziano, R. Talamini, D. Piccolo, A. Gatti, G. Trevisan, G. Sasso, A. Veronesi, A. Carbone, and H. Peter Soyer, "Dermoscopic criteria for melanoma in situ are similar to those for early invasive melanoma," *Cancer*, vol. 91, no. 5, pp. 992–997, 2001.

[2] A. Breslow, "Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma." *Annals of Surgery*, vol. 172, no. 5, pp. 902–908, 1970.

[3] M. Amouroux and W. Blondel, "Non-invasive determination of Breslow index," in *Current management of malignant melanoma*, M. Y. Cao, Ed. InTech, 2011, pp. 29–44. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00626482

[4] M. Stante, V. De Giorgi, P. Cappugi, B. Giannotti, and P. Carli, "Non-invasive analysis of melanoma thickness by means of dermoscopy: A retrospective study," *Melanoma Research*, vol. 11, no. 2, pp. 147–152, 2001.

[5] L. MB, N. P, and B. V, "Excision margins for primary cutaneous melanoma: Updated pooled analysis of randomized controlled trials," *Archives of Surgery*, vol. 142, no. 9, pp. 885–891, 2007.

[6] M. Brady and D. Coit, "Sentinel lymph node evaluation in melanoma," *Archives of Dermatology*, vol. 133, no. 8, pp. 1014–1020, 1997.

[7] D. Rigel, J. Russak, and R. Friedman, "The evolution of melanoma diagnosis: 25 years beyond the abcds," *CA Cancer Journal for Clinicians*, vol. 60, no. 5, pp. 301–316, 2010.

[8] A. Kardynal and M. Olszewska, "Modern non-invasive diagnostic techniques in the detection of early cutaneous melanoma," *Journal of Dermatological Case Reports*, vol. 8, no. 1, pp. 1–8, 2014.

[9] M. Vestergaard, P. Macaskill, P. Holt, and S. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting," *British Journal of Dermatology*, vol. 159, no. 3, pp. 669–676, 2008.

[10] A. Gulia and C. Massone, "Advances in dermoscopy for detecting melanocytic lesions," *F1000 Medicine Reports*, vol. 4, no. 1, 2012.

[11] P. Carli, V. De Giorgi, D. Palli, V. Giannotti, and B. Giannotti, "Preoperative assessment of melanoma thickness by abcd score of dermatoscopy," *Journal of the American Academy of Dermatology*, vol. 43, no. 3, pp. 459–466, 2000.

[12] H. Haenssle, B. Korpas, C. Hansen-Hagge, T. Buhl, K. Kaune, A. Rosenberger, U. Krueger, M. Schn, and S. Emmert, "Seven-point checklist for dermatoscopy: performance during 10 years of prospective surveillance of patients at increased melanoma risk." *Journal of the American Academy of Dermatology*, vol. 62, no. 5, pp. 785–93, 2010.

[13] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, and M. Delfino, "Clinical and dermatoscopic criteria for the preoperative evaluation of cutaneous melanoma thickness," *Journal of the American Academy of Dermatology*, vol. 40, no. 1, pp. 61–68, 1999.

[14] H. Lorentzen, K. Weismann, and F. Grnhj Larsen, "Dermatoscopic prediction of melanoma thickness using latent trait analysis and likelihood ratios," *Acta Dermato-Venereologica*, vol. 81, no. 1, pp. 38–41, 2001.

[15] G. Argenziano, C. Longo, A. Cameron, S. Cavicchini, J.-Y. Gourhant, A. Lallas, I. McColl, C. Rosendahl, L. Thomas, D. Tiodorovic-Zivkovic, P. Zaballos, and I. Zalaudek, "Blue-black rule: a simple dermoscopic clue to recognize pigmented nodular melanoma," *British Journal of Dermatology*, vol. 165, no. 6, pp. 1251–1255, 2011.

[16] V. da Silva, J. Ikino, M. Sens, D. Nunes, and G. Di Giunta, "Dermoscopic features of thin melanomas: A comparative study of melanoma in situ and invasive melanomas smaller than or equal to 1mm [carac-

tersticas dermatoscpicas de melanomas finos: Estudo comparativo entre melanomas in situ e melanomas invasivos menores ou iguais a 1mm],” *Anais Brasileiros de Dermatologia*, vol. 88, no. 5, pp. 712–717, 2013.

[17] I. Maglogiannis and C. N. Doukas, “Overview of advanced computer vision systems for skin lesions characterization,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721–733, 2009.

[18] R. Garnavi, M. Aldeen, and J. Bailey, “Computer-aided diagnosis of melanoma using border- and wavelet-based texture analysis,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1239–1252, 2012.

[19] M. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker, and R. Moss, “A methodological approach to the classification of dermoscopy images,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362–373, 2007.

[20] P. Rubegni, G. Cevenini, P. Sbano, M. Burroni, I. Zalaudek, M. Risulo, G. Dell’Eva, N. Nami, A. Martino, and M. Fimiani, “Evaluation of cutaneous melanoma thickness by digital dermoscopy analysis: a retrospective study,” *Melanoma research*, vol. 20, no. 3, pp. 212–217, 2010.

[21] G. Argenziano, H. Soyer, and et al., *Interactive atlas of dermoscopy.* Milan: EDRA-Medical Publishing and New Media, 2000.

[22] H.-T. Lin and L. Li, “Reduction from cost-sensitive ordinal ranking to weighted binary classification,” *Neural Computation*, vol. 24, no. 5, pp. 1329–1367, 2012.

[23] B.-Y. Sun, J. Li, D. D. Wu, X.-M. Zhang, and W.-B. Li, “Kernel discriminant learning for ordinal regression,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 6, pp. 906–910, 2010.

[24] A. Sáez, C. Serrano, and B. Acha, “Model-based classification methods of global patterns in dermoscopic images,” *IEEE Transactions on Medical Imaging*, vol. 33, no. 5, pp. 1137–1147, 2014.

[25] A. Sáez, C. S. Mendoza, B. Acha, and C. Serrano, “Development and evaluation of perceptually adapted colour gradients.” *IET Image Processing*, vol. 7, no. 4, p. 355 363, 2013.

[26] H. Soyer, g. argenziano, R. Hofmann-Wellenhof, and R. Johr, *Color Atlas of Melanocytic Lesions of the Skin*. Springer Berlin Heidelberg, 2010.

[27] K. Weismann and H. F. Lorentzen, “Dermoscopic color perspective,” *Archives of Dermatology*, vol. 142, no. 9, p. 1250, 2006.

[28] S. Seidenari, G. Pellacani, and C. Grana, “Computer description of colours in dermoscopic melanocytic lesion images reproducing clinical assessment,” *British Journal of Dermatology*, vol. 149, no. 3, pp. 523–529, 2003.

[29] N. Otsu, “Threshold selection method from gray-level histograms.” *IEEE Trans Syst Man Cybern*, vol. SMC-9, no. 1, pp. 62–66, 1979, cited By 10522.

[30] M. Sadeghi, M. Razmara, T. Lee, and M. Atkins, “A novel method for detection of pigment network in dermoscopic images using graphs,” *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 137–143, 2011.

[31] C. Barata, J. Marques, and T. Mendona, “Bag-of-features classification model for the diagnose of melanoma in dermoscopy images using color and texture descriptors,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7950 LNCS, pp. 547–555, 2013.

[32] M. Rastgoo, R. Garcia, O. Morel, and F. Marzani, “Automatic differentiation of melanoma from dysplastic nevi,” *Computerized Medical Imaging and Graphics*, vol. 43, pp. 44–52, 2015.

[33] R. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. smc 3, no. 6, pp. 610–621, 1973.

[34] L.-K. Soh and C. Tsatsoulis, “Texture analysis of sar sea ice imagery using gray level co-occurrence matrices,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2 I, pp. 780–795, 1999.

[35] D. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization,” *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.

[36] B. Manjunath and R. Chellappa, “Unsupervised texture segmentation using markov random field models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 5, pp. 478–482, 1991.

[37] T. Ojala, M. Pietikinen, and T. Menp, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[38] M. Zortea, T. Schopf, K. Thon, M. Geilhufe, K. Hindberg, H. Kirchesch, K. Mllersen, J. Schulz, S. Skrvseth, and F. Godtliebsen, “Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists,” *Artificial Intelligence in Medicine*, vol. 60, no. 1, pp. 13–26, 2014.

[39] P. Cavalcanti and J. Scharcanski, “Texture Information in Melanocytic Skin Lesion Analysis Based on Standard Camera Images,” in *Computer Vision Techniques for the Diagnosis of Skin Cancer*, J. Scharcanski and M. E. Celebi, Eds. Springer Berlin Heidelberg, 2014, pp. 221–242.

[40] J. C. Fernández-Caballero, F. J. Martínez-Estudillo, C. Hervás-Martínez, and P. A. Gutiérrez, “Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks,” *IEEE Transactions on Neural Networks*, vol. 21, no. 5, pp. 750 –770, may 2010.

[41] M. Cruz-Ramírez, C. Hervás-Martínez, J. Sánchez-Monedero, and P. Gutiérrez, “Metrics to guide a multi-objective evolutionary algorithm for ordinal classification,” *Neurocomputing*, vol. 135, pp. 21–31, July 2014.

[42] C. Hervás-Martínez and F. Martínez-Estudillo, “Logistic regression using covariates obtained by product-unit neural network models,” *Pattern Recognition*, vol. 40, no. 1, pp. 52–64, 2007.

[43] C. Hervás-Martínez, F. J. Martínez-Estudillo, and M. Carbonero-Ruz, “Multilogistic regression by means of evolutionary product-unit neural networks,” *Neural Networks*, vol. 21, no. 7, pp. 951–961, 2008.

[44] R. Durbin and D. Rumelhart, “Products units: A computationally powerful and biologically plausible extension to backpropagation networks,” *Neural Computation*, vol. 1, no. 1, pp. 133–142, 1989.

[45] P. A. Gutiérrez, C. Hervás-Martínez, and F. J. Martínez-Estudillo, “Logistic regression by means of evolutionary radial basis function neural networks,” *IEEE Transactionss on Neural Networks.*, vol. 22, no. 2, pp. 246–263, 2011.

[46] N. Landwehr, M. Hall, and E. Frank, “Logistic model trees,” *Machine Learning*, vol. 59, no. 1-2, pp. 161–205, 2005.

[47] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” *Special Interest Group on Knowledge Discovery and Data Mining Explorer Newsletter*, vol. 11, pp. 10–18, November 2009.

[48] P. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F. Fernandez-Navarro, and C. Hervás-Martínez, “Ordinal regression methods: survey and experimental study,” *IEEE Transactions on Knowledge and Data Engineering*, vol. In press, no. 99, 2015.

[49] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[50] W. Chu and S. S. Keerthi, “Support Vector Ordinal Regression,” *Neural Computation*, vol. 19, no. 3, pp. 792–815, 2007.

[51] S. Baccianella, A. Esuli, and F. Sebastiani, “Evaluation measures for ordinal regression,” in *Proceedings of the Ninth International Conference on Intelligent Systems Design and Applications (ISDA’09)*. San Mateo, CA: IEEE Computer Society, 2009, pp. 283–287.